THE ANALYSIS OF CHANGE: INNOVATIONS IN THE VISUAL ANALYSIS OF DATA

Neville M Blampied
Department of Psychology
Seminar, Massey University Albany
05 November, 2015

The Challenge of Change

Change is the heart and soul of Psychology –

- Detect
- Induce
- Measure
- Analyse
- Theorise
- Explain
- Predict & Control

CHANGE

Researching Change

- Focus on within-participant change
- Not between participant difference
  - Lakens (2013)

    “designism”

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, article 863

Analysis of Group data

As soon as you have collected your data, before you compute any statistics, ***look at your data****.*

[Wilkinson & Task Force, 1999, emphasis in original]

So – [Exploratory] Visual analysis

BUT – visual analysis can get unwieldy as N gets larger

E.g., Clyne & Blampied**,** ***Behaviour Change*** | Volume **21** | Number **4** | 2004

Looking at change data

Some questions -

Is there a better way?

Is there an answer to Barlow's question?

Barlow & Nock (2009). Why cant we be more idiographic in our research? *Perspectives on Psychological Science, 4*, 19 – 21.

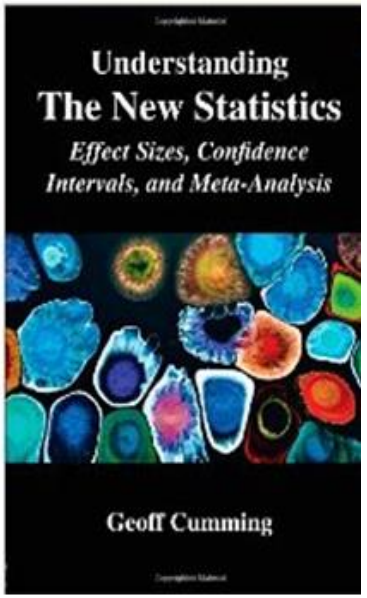
Can we integrate with *the new statistics*?

confidence intervals, effect sizes etc ….

The new statistics

- Estimation }
- Precision – }*
  - Confidence intervals
- Effect sizes
- Meta-analysis
  - (best evidence synthesis)

*Both lead to concern for Measurement: validity/reliability/error



- Does not routinely use NHST

. . . *friends do not let friends compute p* [Klein, 2013].

I conclude from the arguments and evidence I have reviewed that best research practice is not to use NHST at all [Cumming, 2012]

Beginning at the beginning – Brinley Plots - 1965

Beginning at the beginning – Brinley Plots - 1965

Brinley's insight

Scatterplots - If

- X scale = Y scale, &
- Common origin
- Then 45^0 diagonal =

  line of no effect/X = Y

Systematic effects deviate from line

- Coordinate pairs plotted represent sub-group means grouped by categorical variable (age)

Beginning at the beginning – Brinley Plots - 1965

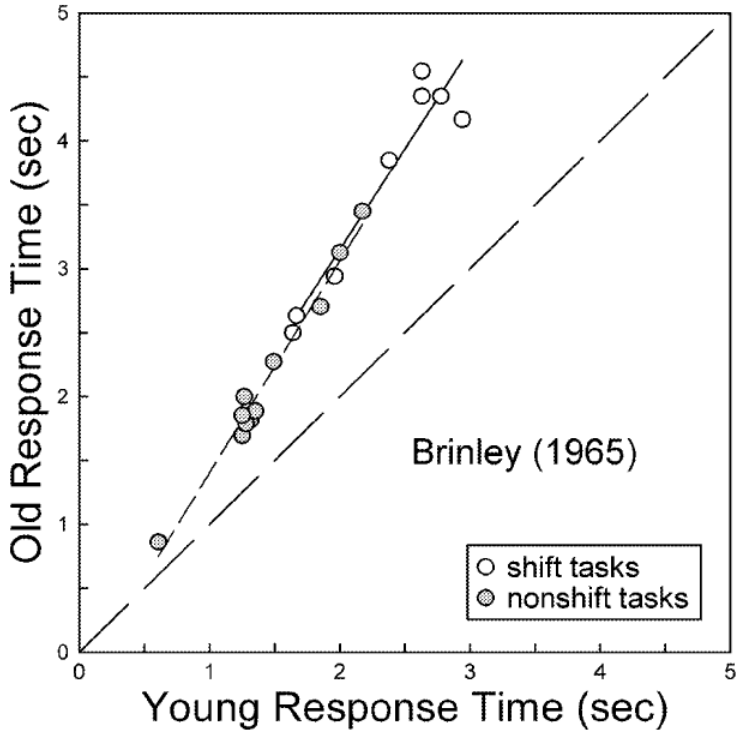


Figure 1. Mean response time (RT) of the older adult group as a function of the mean RT of the young adult group in the corresponding experimental condition. The solid line is fit to the data from the 9 shift conditions (open circles) and the dashed line is fit to the data from the 12 nonshift conditions. If the condition mean RTs for the old and young groups were equal, the points would fall along the diagonal. Data are taken from Brinley (1965).

Brinley, J.F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In A.T. Welford & J.E. Birren (Eds.). *Behavior, ageing, and the nervous system* (pp 114 – 149). Springfield, IL: Charles C. Thomas.

Fig 1 from Myerson, J, Adams, D.R., Hale, S., and Jenkins, L. (2003). Analysis of group differences in processing speed: Brinley plots, Q-Q plots, and other conspiracies. *Psychonomic Bulletin & Review, 10*, 224 – 237.

Contemporary Example: Dye, Green, & Bavelier, (2009). Increasing speed of processing with action video games. ***Current Directions in Psychological Science***,18 (6), p321-326

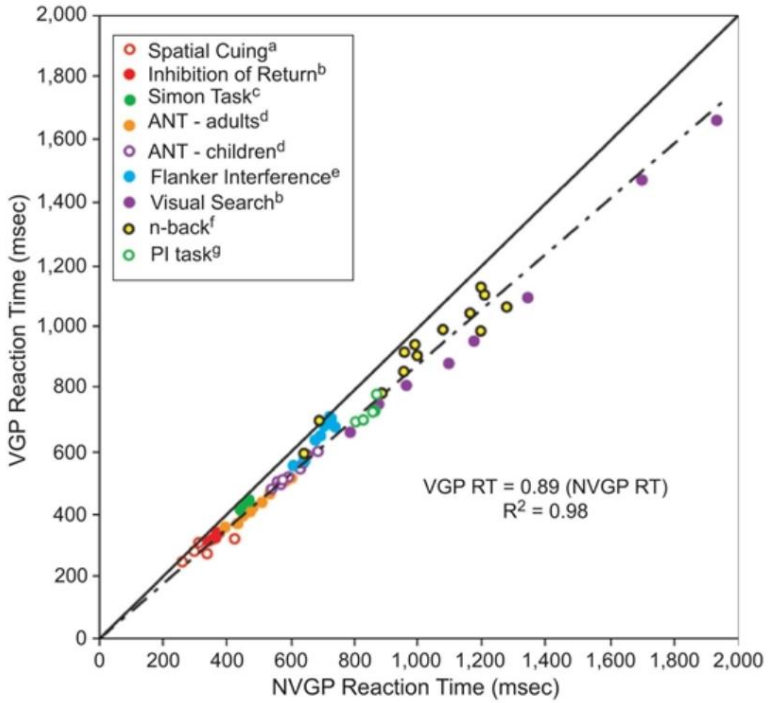
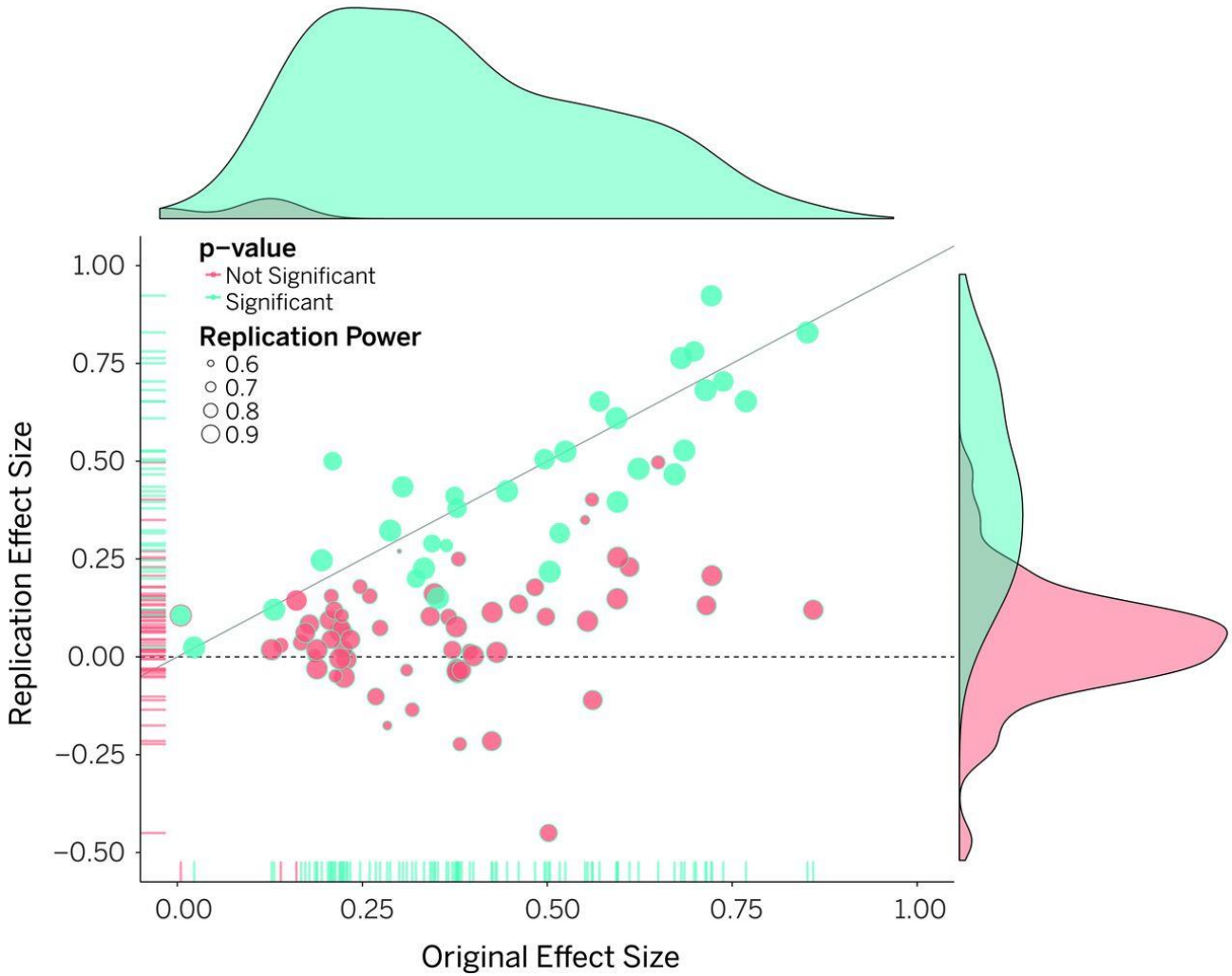


Fig. 1. A Brinley plot showing the reaction time (RT) of non-video-game players (NVGPs) on the X-axis versus that of expert video-game players (VGPs) on the Y-axis, for 89 different experimental conditions from nine different types of task. For each experimental condition, the RTs of VGPs and NVGPs were retrieved and plotted as one separate data point. A simple linear function ($y = mx$) was used to describe the relationship between VGP and NVGP RTs (dashed line). VGPs responded 11% faster than NVGPs across a wide range of RTs (VGP RTs = .89 × NVGP RTs, $R^2 = 0.98$). Importantly, similar accuracy was observed across groups, ruling out an explanation in terms of simple speed-accuracy trade-off (VGP accuracy = 0.99 × NVGP accuracy, $R^2 = 0.92$). The studies are (a) Greenfield, deWinstanley, Kilpatrick, & Kaye (1994); (b) Castel, Pratt, & Drummond (2005); (c) Bialystok (2006); (d) Dye, Green, & Bavelier (2009); (e) Green & Bavelier (2003); (f & g) Bavelier & Bailey (2007).

Original study effect size versus replication effect size (correlation coefficients).



Open Science Collaboration Science 2015;349:aac4716
http://www.sciencemag.org/content/349/6251/aac4716

First use in clinical context - 1979

Brinley plot modified

- **Individual's** data points
  @ t_1 (X-axis) plotted against t_2 (Y-axis)

- t_1 = weight change @ end of treatment

- t_2 = weight change @ 5 yr follow-up

- Stable weight = points on the line

Stunkard, A.J., & Penick, S.B. (1979). Behavior modification in the treatment of obesity. *Archives of General Psychiatry, 36*, 801 – 806.

First use in clinical context - 1979

Brinley plot modified

- **Individual's** data points
  @ t_1 (X-axis) plotted against t_2 (Y-axis)
- t_1 = weight change @ end of treatment
- t_2 = weight change @ 5 yr follow-up
- Stable weight = points on the line

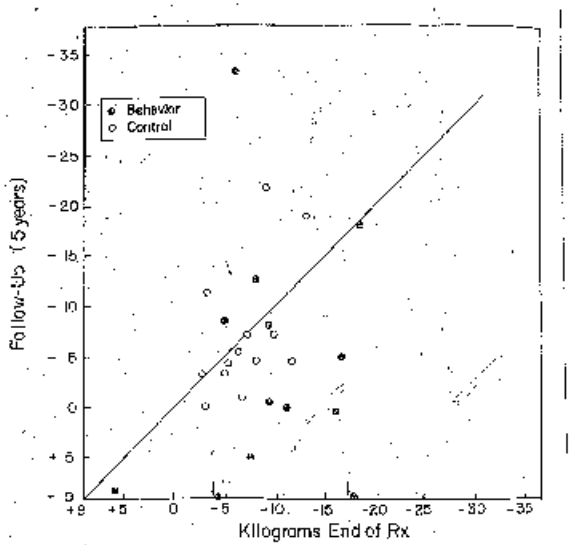


Fig 2.—Weight changes from end of treatment to five-year follow-up.

Stunkard, A.J., & Penick, S.B. (1979). Behavior modification in the treatment of obesity. *Archives of General Psychiatry, 36*, 801 – 806.

Other clinical examples

Therapy for alcohol abuse

- Sobell, M.B., Sobell, L.C., & Gavin, D.R. (1995). Portraying alcohol treatment outcomes: Different yardsticks of success. *Behavior Therapy, 26*, 643 – 669.

CBT for marital distress

- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336 – 352.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting & Clinical Psychology, 59*, 12 – 19.

Other clinical examples

Therapy for alcohol abuse

- Sobell, M.B., Sobell, L.C., & Gavin, D.R. (1995). Portraying alcohol treatment outcomes: Different yardsticks of success. *Behavior Therapy, 26*, 643 – 669.

Now regarded as the *classical* approach →

CBT for marital distress

- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336 – 352.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting & Clinical Psychology, 59*, 12 – 19.

Other clinical examples

Therapy for alcohol abuse

- Sobell, M.B., Sobell, L.C., & Gavin, D.R. (1995). Portraying alcohol treatment outcomes: Different yardsticks of success. *Behavior Therapy, 26*, 643 – 669.

BUT, these innovations had almost no impact on subsequent clinical research!

CBT for marital distress

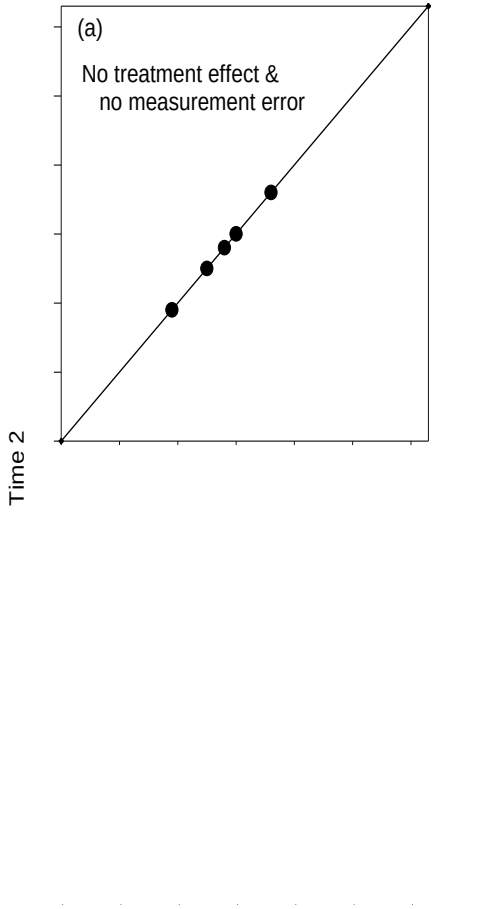
- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336 – 352.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting & Clinical Psychology, 59*, 12 – 19.

I'm following & extending this work

Modified Brinley plots – key features

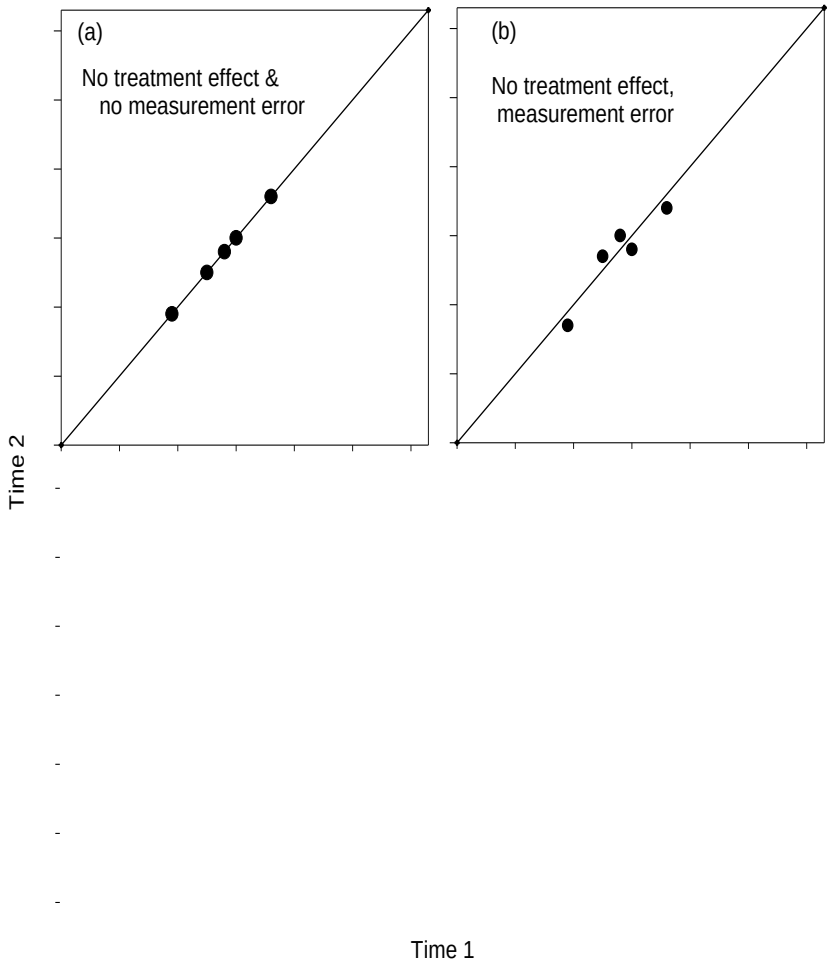
- Individual's data is plotted
- Axes same scale & origin
- 45º diagonal is line of no effect

(a) No time effect shown – perfect stability x = y



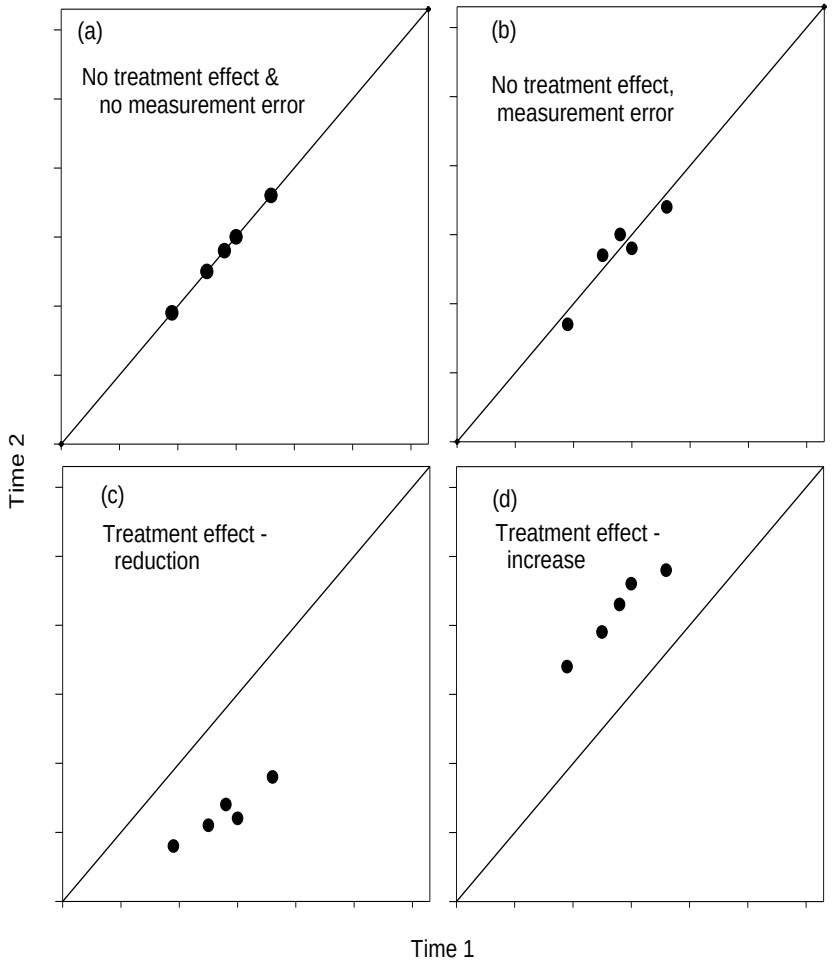
Modified Brinley plots – key features

(b) Unsystematic variability/measurement error



Modified Brinley plots – key features

(c & d) Systematic change over time is shown as points above/below the line



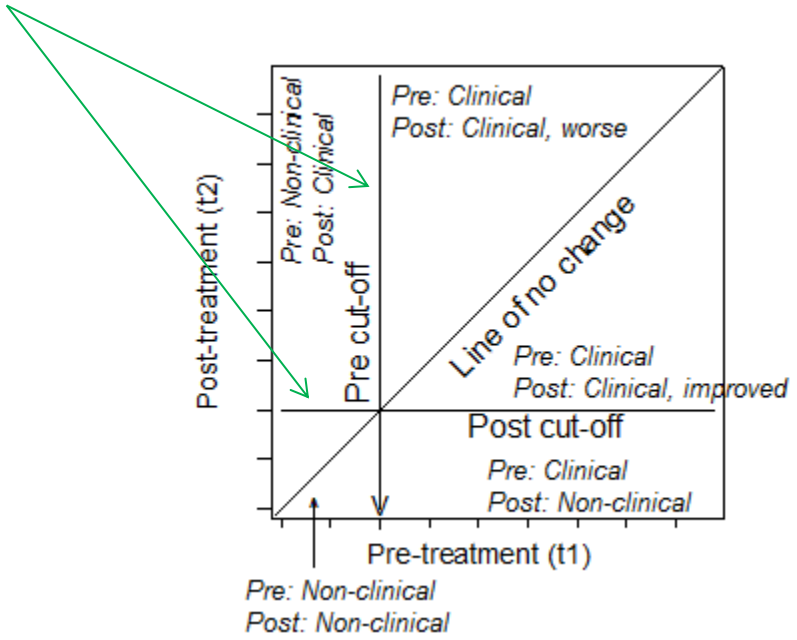
Modified Brinley plots – aids to interpretation

Add clinical cut-off lines

- After Jacobson, et al

Modified Brinley plots – aids to interpretation

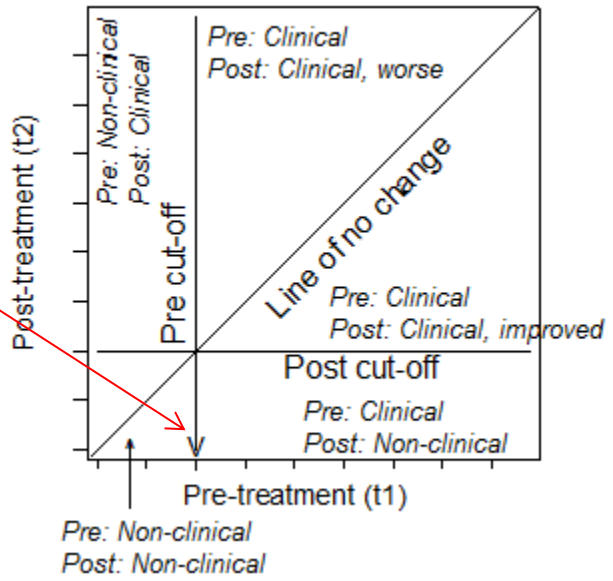
Clinical cut-off lines



Modified Brinley plots – aids to interpretation

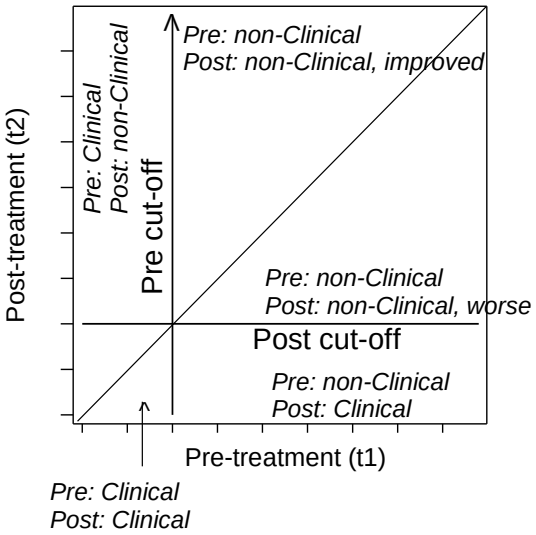
arrow indicates score reduction = improvement

- Graph sectors have meaningful interpretation



Aids to interpretation …

Interpretation where increase = clinical improvement



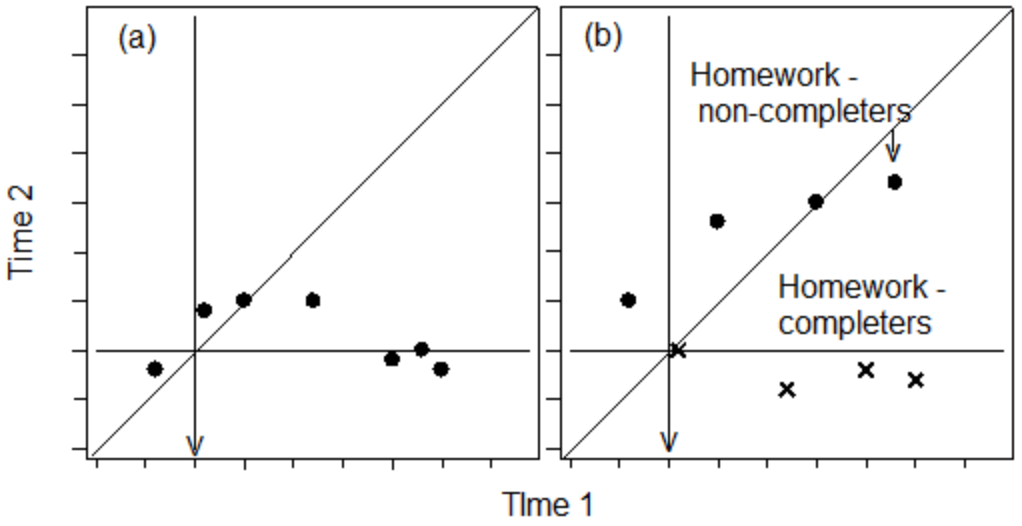
Further refinements - Examples

(a) Shows possible interaction between treatment and initial severity of problem

(b) Shows effect of categorical variable split.

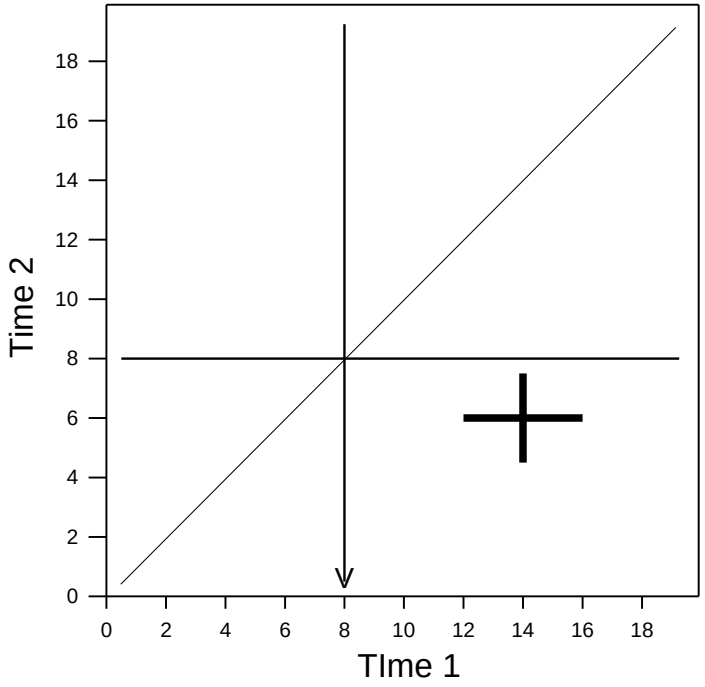
Many categorical variables can be investigated –

- gender
- age
- ethnicity
- therapist
- therapy features
- etc



Interpreting the plot …
Group information

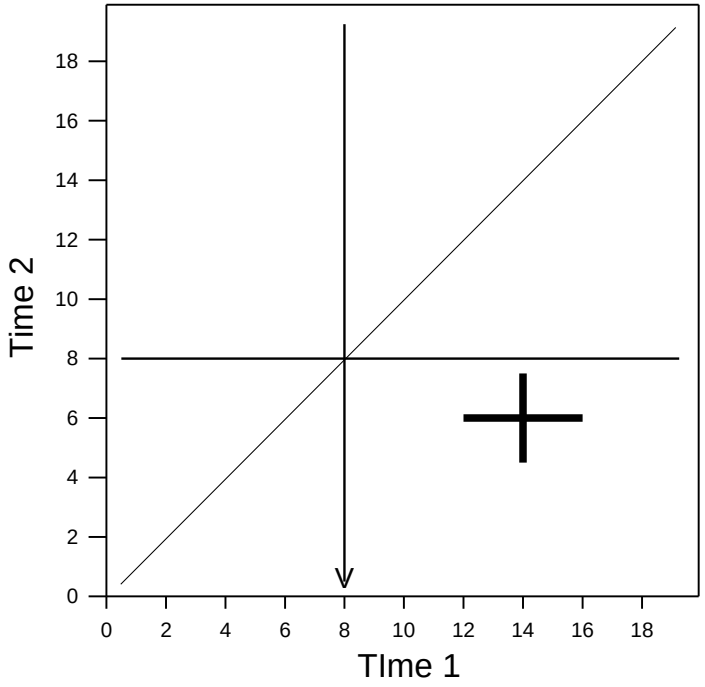
Means @ t_1 & t_2



+ marks the means

Interpreting the plot … Confidence Intervals

Means + 95% CI



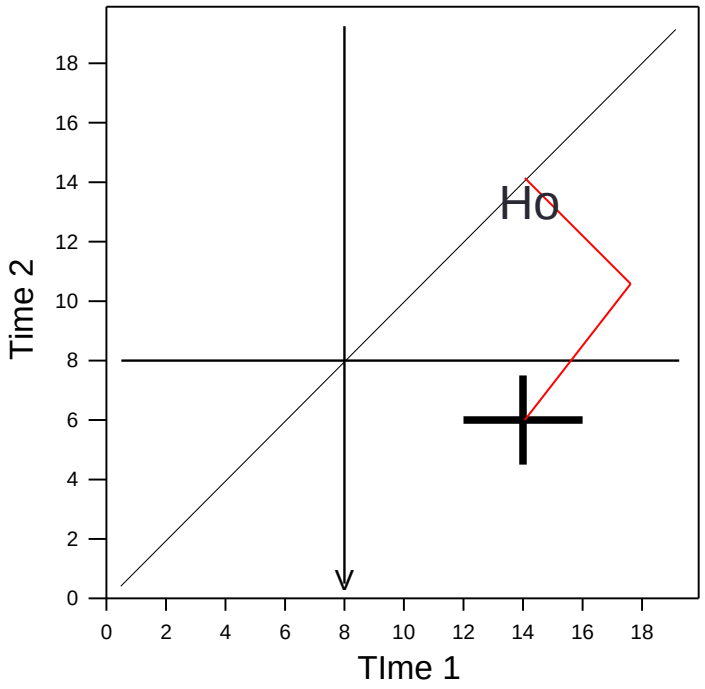
What does the CI mean?

The interval [95% CI] estimates μ, with 95% confidence.

[Klein, R.B. (2013) *Beyond significance testing* 2nd Ed. p 41]

Interpreting the plot … *t*-test

t-test



Null hypothesis is that the point lies on the diagonal

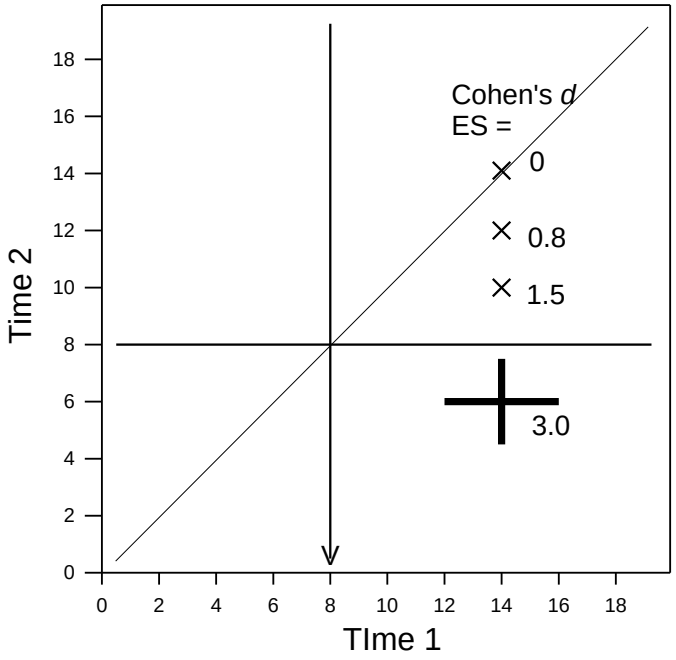
Interpreting the plot … Cohen's *d*

Warning!

Calculation of *d(within)* is complicated.

See Lakens (2013)

Cohen's *d* Effect size



How much change is needed to believe that it is real change?

Suppose

- *d* is large
- *t* is statistically significant

Therefore group mean change is OK

But what about individual change?

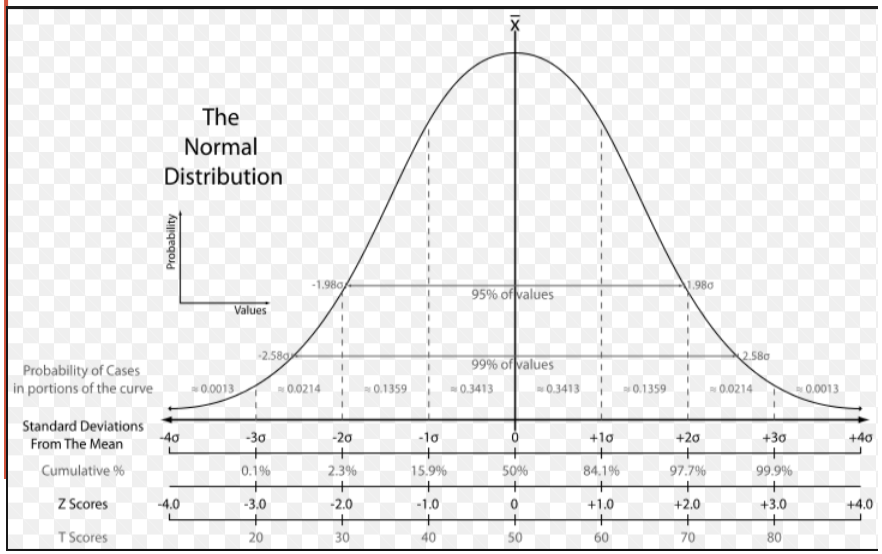
How much is enough to be real?

The Reliable Change Index

(Jacobson & Truax, 1991)

Reliable Change & Measurement Error

Distribution of measurement error



RC – what you need to know to compute

Info about the measure

- S = SD of reference data-set
- r_{xx} = Test-retest reliability of measure (Chronbach's alpha)

Used to compute

1. SEM
2. SDIFF

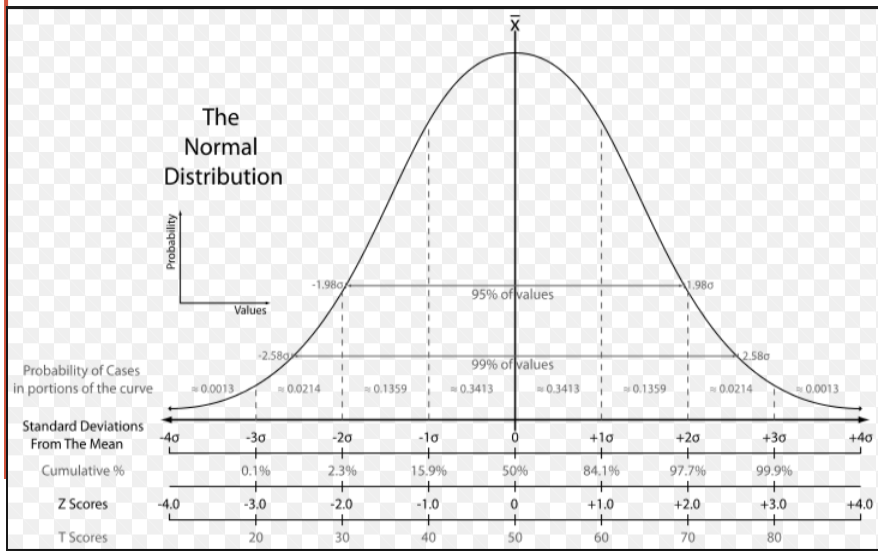
Both are a form of Standard Deviation

SDIFF is SEM of the Error Distribution of the Difference Scores

Distribution of measurement error

- Is a Normal distribution
- SEM/ SDIFF is the Standard Deviation of the error distribution
- 95% of errors lie within

  +/- 1.96 SEM



RC computation

Steps & formulae

1. Compute Standard Error of Measurement

$$SE_M = s\sqrt{1-rxx}$$

2. Compute S_{DIFF}

$$S_{DIFF} = \sqrt{2(SEM^2)}$$

3. Compute the difference score for each individual

$$Diff = x_1 - x_2$$

4. Compute $x_1 - x_2 / S_{DIFF}$

5. If ≥1.96 a difference that large not likely due to measurement error – is in 5% tail of error distribution

Change not likely due to measurement error p < .05

Example

1. If variability s = 7.5

Test reliability r_{xx} = .80

$SE_M = 7.5\sqrt{1-.8} = 3.35$

2. $S_{DIFF} = \sqrt{2(3.35*3.35)} =$ 4.74

3. So if

x1 = 47.75

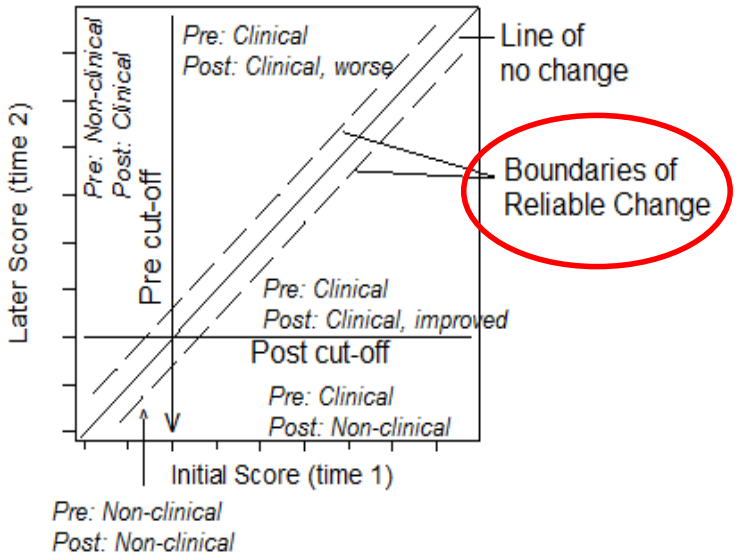
x2 = 32.5 Diff = 15

4. 15/4.74 = 3.16

3.16 > 1.96 – **Change is reliable**

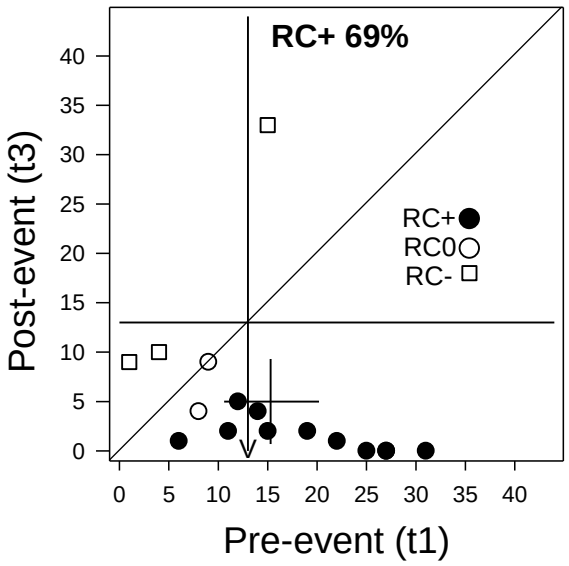
Displaying RC information

By lines



Reliable Change Index (boundaries) = $\pm S_{Diff}$ x 1.96

Displaying RC information …



By classifying participants RC+, RCo, RC-

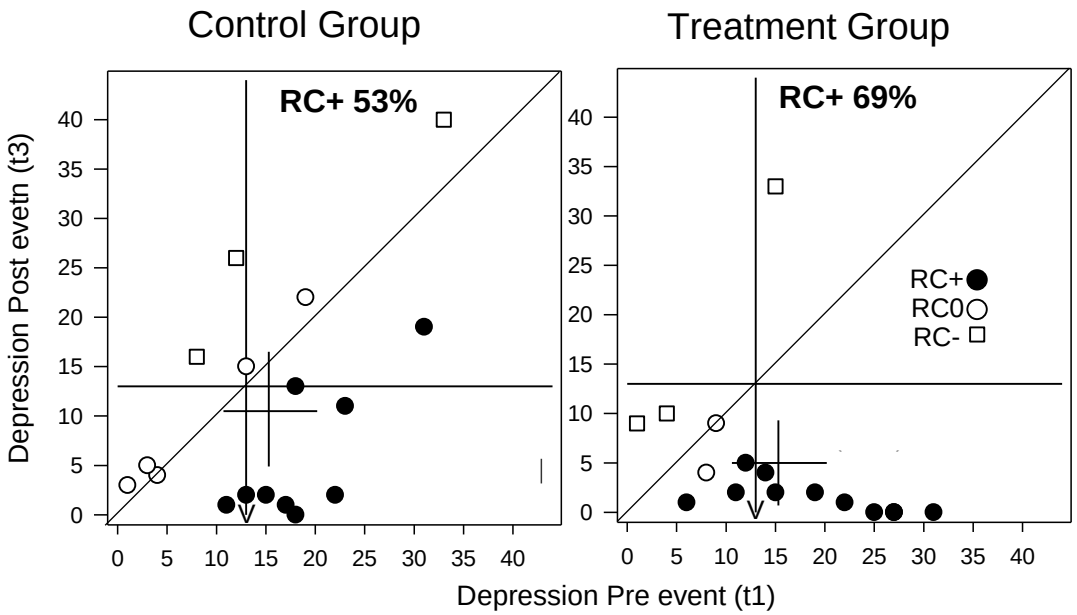
Time 1	Time2	t1 - t2	RC
31	0	31	+
27	0	27	+
27	0	27	+
25	0	25	+
22	1	21	+
19	2	17	+
15	2	13	+
14	4	10	+
11	2	9	+
12	5	7	+
6	1	5	+
9	9	0	o
7	4	3	o
4	10	-6	-
1	9	-8	-
15	33	-18	-

RC example – between groups

RC+% = RC+/n x 100

Is another Effect Size measure

- Measures individual impact
- Vs
- Cohen's *d*
- Measures group mean impact



(data from Rucklidge & Blampied, 2011).

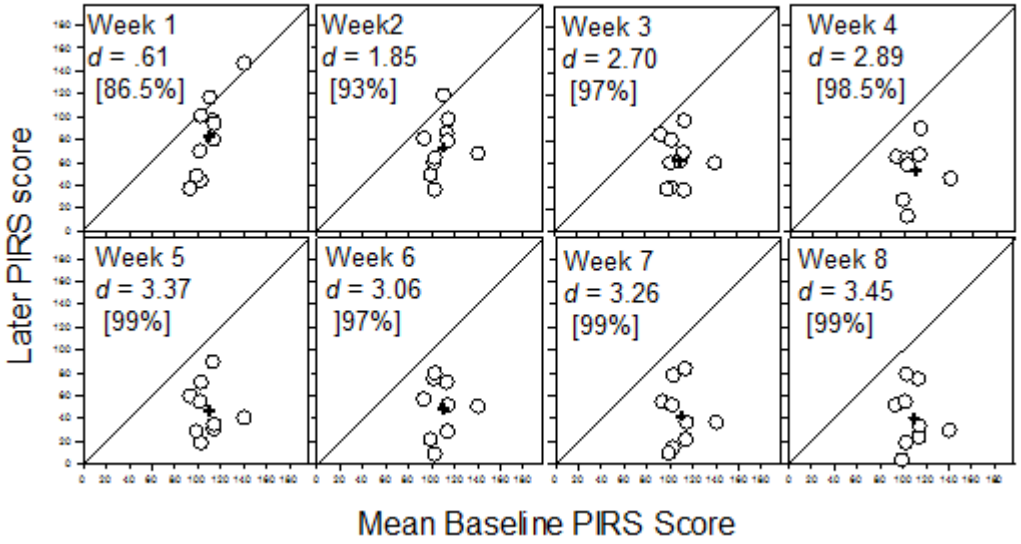
Tracking change over time – plus two Effect size measures

- Group mean +
- ES1 = d_{rm}
- *ES2* =

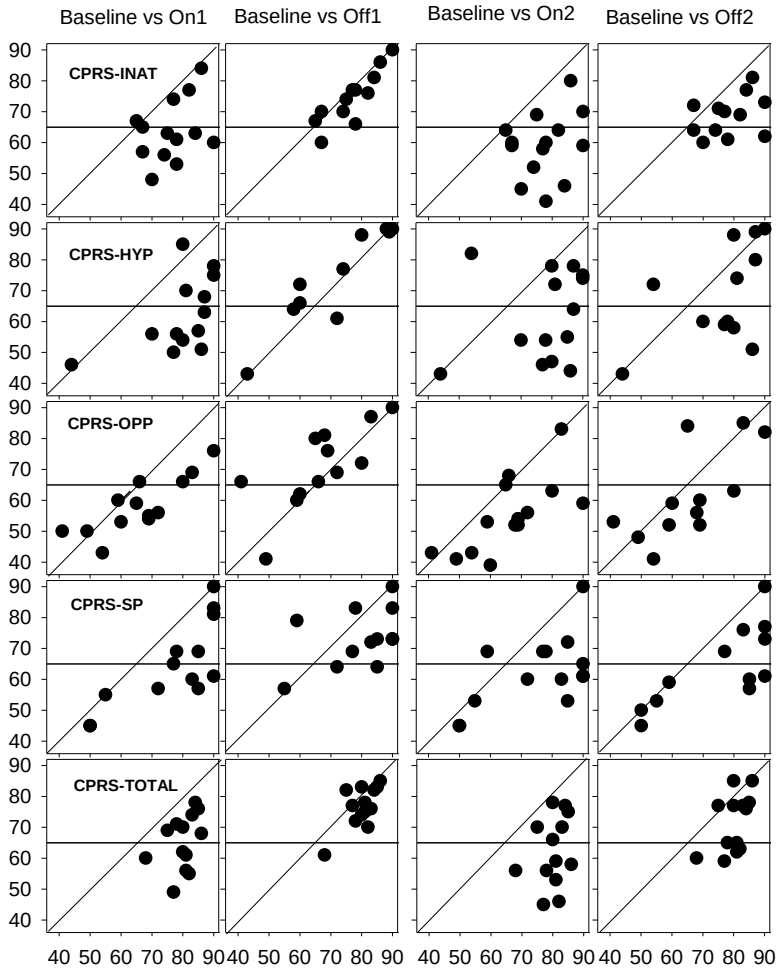
  [Common Language Effect Size] =

The probability of having a better score at time 2 than time 1

(data from Lothian, 2014)

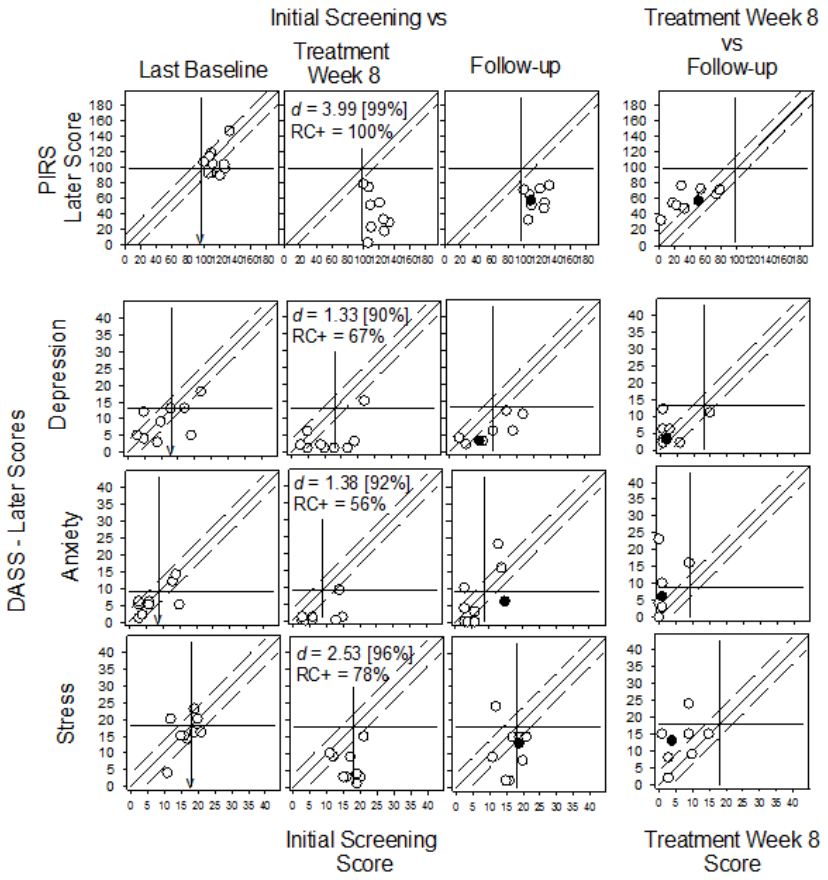


Summarising much data (Gordon, 2014)



14 subjects x 5 phases x 5 DVs
= 350 data points

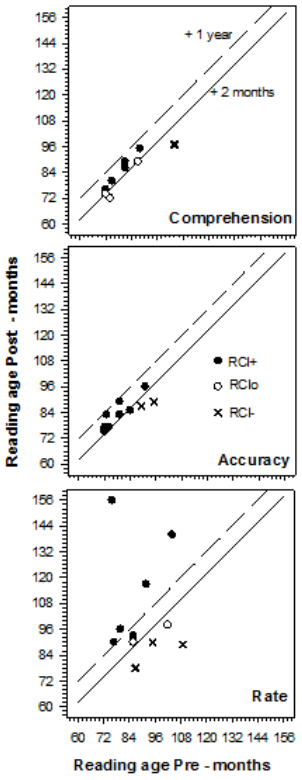
Confirming that baselines are stable



& looking at follow-up relative to post-treatment

[Lothian, 2015]

Adapting for absolute changes over time



- Solid diagonal shifted upwards represents

Y = X(age) + 2mo

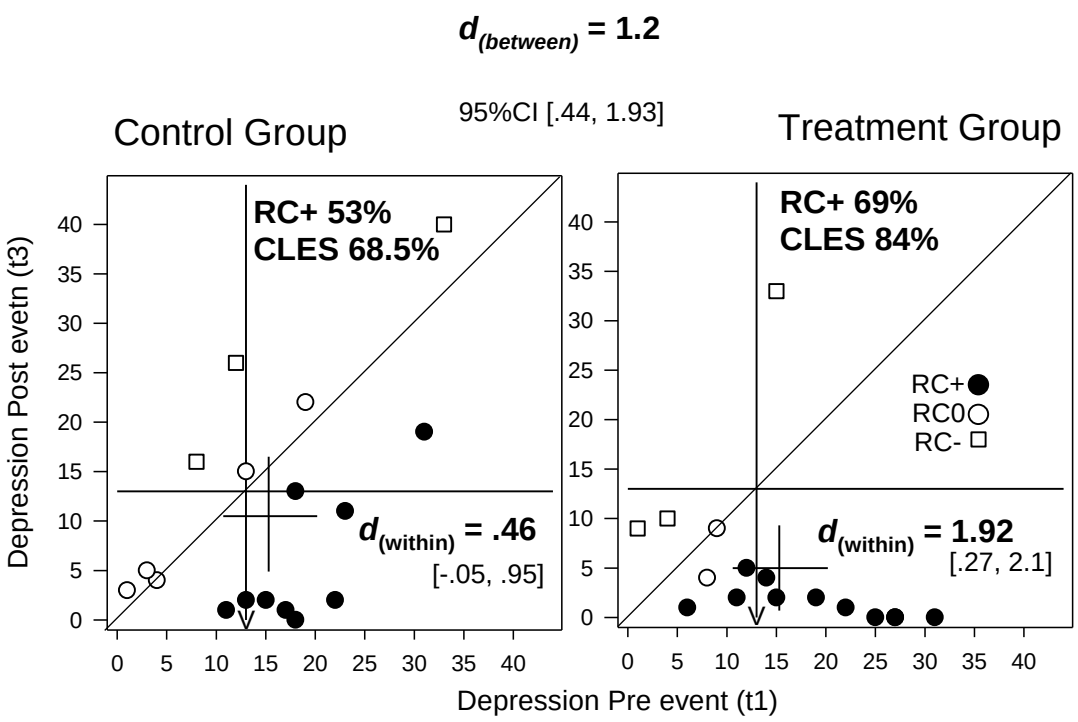
(Reading age)

- Dashed diagonal represents

Y = X(age) + 12mo

Robson, Blampied & Walker (2015)

One graph to rule them all? – modified Brinley Plots



- Visual analysis √
- Groups √
- Means √
  - 95% Confidence intervals √
- Individuals √
- Reliable Change √

  (individuals p <.05)

Effect size √

- 95% CI on ES √
- % with Reliable Change √
- Clinical significance √

Potential for a new methodological synthesis for merging nomothetic & idiographic research?

nomothetic

- Concerned with general laws
- Concerned with the universal
- Abstract
- Timeless
- Objective/impersonal
- Inter-individual research
- Legacy of Quetelet/Fisher

idiographic

- Concerned with the individual case in context
- Concerned with the particular
- Concrete
- Historically situated
- Subjective/personal
- Intra-individual research
- Legacy of Bernard/Pavlov/ Skinner

Potential synthesis merging nomothetic & idiographic research & the new statistics?

The new statistics$_{(nomothetic)}$ + single-case/ individuals$_{(idiographic)}$ + replication$_{(both)}$ + visual analysis$_{(both)}$

Abandon our over-reliance on NHST *p <.05*

And

- Show proper respect for measurement – Calibration & SEM
- Use Reliable Change @ individual level
- Attend to clinical/practical significance [Effect size] instead of statistical significance

Humans, not the gods, created all forms of enquiry, and we can and should modify them.
[Camic, Rhodes, & Yardley, 2003, p4]

Acknowledgements & Thanks

Colleagues

Julia Rucklidge

Karyn France

Jacki Henderson

Lawrence Walker

Mark Sobell

Bill Fordyce

Students – for letting me play with their data

Jo Capstick

Amy Clarke

Heather Gordon

Martin Hood

Wendy Kelly (VUW)

James Maguire

Jess McIvor (Massey)

Joanna Lothian

Pip Newton

Cathy Robson

Ellen Sole

Richard Straight

Vic Weedon (Massey)

Abstract

Methodological reform in psychology calls for research to be more idiographic and less dependent on group statistical inference. Recommended alternatives include more extensive use of graphs and visual analysis. This paper describes the construction and interpretation of modified Brinley plots, a technique for analysing treatment outcomes for individuals within groups that is particularly suitable for outcome research of psychological therapies. Modified Brinley plots are scatter-plots that compare individual scores at time 1 (normally pre-treatment) with scores at various times post-treatment. If the origin and axis scales are the same no or little change is shown by data points clustering on or about the 45^{o} diagonal line. Change over time (improvement or deterioration) is shown by shifts away from the diagonal. Interpretation is aided by the addition of clinical cut-offs, and by the use of the Reliable Change Index (based on measurement error). In addition to displaying individuals' data, these graphs may also display group effects such as means, variances, confidence intervals, and effect sizes. Both between-group and within-group data may be presented and analysed this way and large amounts of data can be efficiently presented and clearly understood within one figure. **This talk may be particularly helpful to students planning research into within-participant change over time**